JOURNAL OF ENVIRONMENTAL ECONOMICS AND MANAGEMENT **33,** 75–93 (1997)
ARTICLE NO. EE960981

# Evaluating the Performance of Benefit Transfer: An Empirical Inquiry

Stefanie Kirchhoff,* Bonnie G. Colby,† and Jeffrey T. LaFrance†

**Department of Agricultural Economics, University of Maryland, Baltimore, Maryland; and*
*† Department of Agricultural and Resource Economics, P.O. Box 210023, Economics Building,*
*University of Arizona, Tucson, Arizona 85721-0023*

Received December 31, 1994; revised May 1, 1996

Benefit transfers are used by public agencies needing information on costs and benefits of policy decisions, although scientific debate regarding the validity of benefit transfer is ongoing. This article develops a methodology to evaluate the performance of direct benefit transfer and benefit function transfer and applies the methodology to two pairs of similar non-market amenities. Empirical results indicate that benefit function transfer is more robust than transfer of average site benefits. Our results suggest that the circumstances under which benefit function transfer provides valid, policy-relevant information may be limited and that errors from applying benefit transfer can be quite large, even across seemingly similar amenities.

## INTRODUCTION

Benefit–cost analysis can assist policy makers in making decisions about the preservation of natural environments and allocation of natural resources among alternative uses. In the United States, government agencies whose actions fall under Executive Order 12291 must conduct a benefit–cost analysis for all major regulations [10]. Budget constraints prevent public agencies from conducting an original benefit estimate study for every site that may be affected by proposed regulations and so public agencies faced with growing demands for valuation of non-market amenities are inclined to use benefit transfer, to the extent it provides valid data for policy analysis. Boyle and Bergstrom [2] define ‘‘benefit transfer’’ as ‘‘the transfer of existing estimates of non-market values to a new study which is different from the study for which the values were originally estimated’’ (p. 651). The site for which the original estimates were obtained is often referred to as the ‘‘study site,’’ while the site under consideration for a new policy is termed the ‘‘policy site’’ [5].

Although benefit transfers currently are used in decision making by public agencies, the scientific debate over benefit transfer continues and many issues remain unresolved. In what circumstances do existing benefit estimates provide a credible basis for policy decisions involving sites other than the study site for which benefits were estimated? Study site benefit estimates depend on the demand functions of site visitors. These depend in turn on the specific site attributes as well as socioeconomic characteristics of the recreationists, their preferences, and the price and availability of substitute sites. None of these factors will be identical across different sites and economists would expect, *a priori*, that study site benefits

0095-0696/97 $25.00

will differ from policy site benefits. Given the fact that policy makers are using benefit transfer, it is important to improve our understanding of the conditions under which this approach can provide a reasonable approximation of the true policy site benefits.

This article presents the results of a concurrent estimation of compensating variation for two pairs of water-dependent recreation sites, which (within each pair) are similar with respect to location and recreational activities. Benefit estimates and bid functions are developed and then used to evaluate the performance of direct benefit transfer and benefit function transfer for each pair of sites. Alternative measures of performance are applied to evaluate the reliability of benefit transfer, examining both the transfer of an estimated bid function to a different site and the transfer of mean WTP values from one site to another. The article summarizes the results of testing three alternative hypotheses regarding the validity of benefit transfer procedures and discusses their implications for the practice of benefit transfer. The hypothesis testing procedures reported here may contribute to broader efforts to develop an overall protocol for assessing the validity of benefit transfers [15].

## BACKGROUND

To account for differences in site-specific factors, most authors agree that the transfer of demand functions or value functions (‘‘benefit function transfer’’) is preferred over the direct transfer of average unit values (which we refer to as ‘‘direct benefit transfer’’). Using the coefficient estimates from a study site, demand equations provide a way of accounting for differences in explanatory variables between the study site and the policy site. However, direct benefit transfer often is used where either the benefit function for the study site or the values of the independent variables for the policy site are unavailable. In fact, a large portion of benefit transfer efforts to date have involved transferring mean site benefit measures rather than demand functions or WTP equations [11].

One research approach to evaluate the performance of benefit transfer is the concurrent estimation of non-market values at the study and policy site using primary data collected at both sites. The site-specific benefit estimates are then compared with those derived from benefit transfer [2, 5, 14]. If the two sets of benefits estimates are not statistically different, convergent validity of benefit transfer is presumed.[1] On the other hand, if the estimates obtained by benefit transfer are statistically different from those obtained at the original site, then researchers need to examine the size of the bias, the direction of the bias and methods to adjust the study site estimates to mitigate bias. Using this approach, researchers can identify conditions under which benefit transfer provides a reasonable approximation of policy site benefits and can develop criteria and procedures to enhance the validity of benefit transfer as a policy tool [2]. This approach is based on the premise that study site estimates are a ‘‘true’’ measure of benefits and that deviations from these estimates can be characterized as ‘‘bias.’’ Several prior studies have followed this line of inquiry.

[1] Convergent validity is tested by examining whether CVM estimates (here obtained through benefit transfer) are correlated with alternative measures of the same theoretical construct (here the site-specific estimates) [12].

Loomis [11] tested the performance of travel cost method (TCM) demand equation transfer for recreational fishing in Oregon, Washington, and Idaho. He estimated identically specified multi-site TCM demand equations and then statistically compared the coefficients, testing the cross-state transferability of demand equations. In addition, he used the demand equation obtained from $n - 1$ Oregon rivers to predict benefits at the $n$th river, thereby determining the percentage error of within-state benefit transfers. Loomis' results led to rejection of the equality of demand coefficients across states. Benefit transfers among rivers within the state of Oregon are accurate to within 5 to 15%. Loomis [11] also compared the performance of demand equation transfer to the technique of simply using the average benefits per trip as an indicator of benefits for the unstudied site. He found that the error margins generally are much higher for the latter practice. Loomis notes several shortcomings of his study. Most importantly, the data sets for differing sites were collected at different dates and the data was collected for purposes other than demand estimation, resulting in a relatively simple demand specification limited by the weakest of the multiple data sets.

Downing and Ozuna [6] tested the reliability of the transfer of value functions (benefit function transfer) using dichotomous choice CVM data collected from anglers surveyed across eight contiguous Texas Gulf Coast regions over three distinct time periods. The authors use a total of 128 regressions to analyze the reliability of benefit transfer across time periods, as well as across sites and time periods. They found that the equality of the regression coefficients for study site and policy site could not be rejected in many cases. However, when testing the statistical equality of the welfare measures using 95% confidence intervals, the authors found that the welfare measures were statistically different in about 90% of the cases. They conclude that benefits function transfer is not a reliable approach for the case at hand. Moreover, they stress that testing the equality of regression coefficients is not sufficient to ensure the reliability of benefit function transfer, due to the non-linearity of the logit model used to estimate benefit functions and non-linearity of the benefit estimates themselves.

Opaluch and Mazzotta [14] stress the importance of additional multi-site studies to provide a broader basis for evaluating the adequacy of benefit transfer and formulating standards for acceptability in different contexts. The two studies just described were limited with respect to the number of explanatory variables included in the valuation functions used for benefit transfer. Downing and Ozuna, for instance, include only the bid amounts as explanatory variables and so do not consider socio-economic characteristics of the recreationists. The research reported here presents a comprehensive procedure to test the performance of benefit transfer in the framework of a Tobit model. The Tobit model is preferable to the more commonly used OLS regressions in estimating bid functions from non-dichotomous choice surveys because WTP is a limited dependent variable. However, relatively few studies have applied the Tobit model in a non-dichotomous choice CVM framework [7].

## THE MODEL

Consider the recreating consumer's decision to go or not to go on a single trip to a specific recreation site. Assuming a standard random utility framework, the individual will choose to take the trip if and only if

$$u(m,0,\mathbf{q},\mathbf{d},\varepsilon) \leq u(m-p,1,\mathbf{q},\mathbf{d},\varepsilon), \tag{1}$$

where $u(\cdot)$ is the consumer's utility function, $m$ is total consumption expenditures on goods other than the recreation trip, $t \in \{0,1\}$ is an indicator variable for whether or not the person takes the trip, with $t = 1$ indicating yes and $t = 0$ indicating no, $p$ is the total cost of the recreation trip, $\mathbf{q}$ is a vector of parameters describing the quality of the recreational site/experience, $\mathbf{d}$ is a vector of consumer characteristics observable to the analyst, and $\varepsilon$ is a stochastic component that is known to the consumer but unobservable to the analyst.

The critical trip cost, say $\tilde{p}(m,\mathbf{q},\mathbf{d},\varepsilon)$, at which the individual is just indifferent between taking the trip and not taking the trip is defined by

$$u(m,0,\mathbf{q},\mathbf{d},\varepsilon) \equiv u(m-\tilde{p}(m,\mathbf{q},\mathbf{d},\varepsilon),1,\mathbf{q},\mathbf{d},\varepsilon), \tag{2}$$

This value is the consumers' *total willingness to pay* (TWP) for the recreational experience under consideration. All actual trip costs satisfying $p \leq \tilde{p}(m,\mathbf{q},\mathbf{d},\varepsilon)$ lead the consumer to take the trip, while all trip costs satisfying $p > \tilde{p}(m,\mathbf{q},\mathbf{d},\varepsilon)$ lead the consumer to not take the trip. Thus, the *compensating variation*, $cv(m,p,\mathbf{q},\mathbf{d},\varepsilon)$, for taking this trip at actual cost $p$ is defined by

$$cv(m,p,\mathbf{q},\mathbf{d},\varepsilon) \equiv \tilde{p}(m,\mathbf{q},\mathbf{d},\varepsilon) - p. \tag{3}$$

There are three important aspects of this formulation of the random utility recreational choice model. First, the participating consumer's TWP does not depend explicitly on trip cost. Therefore, when compensating variation is the dependent variable, the coefficient on $p$ is known *a priori* to equal $-1$. This parametric restriction should be incorporated into the regression model to increase the efficiency of the remaining parameter estimates. Second, TWP must equal or exceed actual trip expenditures for participants. Therefore, compensating variation is truncated from below at zero. Thus, a Tobit censored regression framework is necessary to capture this influence, especially when there are valid bid responses of zero compensating variation. This is in fact the case in all four of the recreational sites and activities considered in this study. Third, by definition TWP satisfies

$$\tilde{p}(m,\mathbf{q},\mathbf{d},\varepsilon) \equiv cv(m,p,\mathbf{q},\mathbf{d},\varepsilon) + p. \tag{4}$$

That is, total willingness to pay for any recreational experience is equal to the sum of the cost and the compensating variation for that experience. Hence, an observationally equivalent approach is to add trip expenditures to the contingent valuation bid responses for compensating variation and estimate TWP as the dependent variable subject to the inequality

$$\tilde{p}(m,\mathbf{q},\mathbf{d},\varepsilon) \geq p \tag{5}$$

for all respondents. We follow this latter approach, using the Tobit estimation procedure with a variable truncation point equal to actual trip expenditures to properly account for the censoring of the dependent variable in the regression model.

## SITES

This study focuses on two pairs of recreation sites, one pair in southern Arizona and one pair in northern New Mexico. Ramsey Canyon Preserve in southern Arizona is a relatively small canyon with riparian features dependent on the perennial flows of Ramsey Creek. It is southern Arizona's most prominent birding site and is nationally recognized for its wide variety of hummingbirds. The San Pedro Riparian National Conservation Area (RNCA), also in southern Arizona, is the most extensive continuous stretch of riparian habitat remaining in the southwest desert region. The San Pedro RNCA also is nationally renowned for bird watching due to the wide variety of bird species which can be found within the area.

The study sites in New Mexico are two subsequent stretches of the Rio Grande River: the Taos Box and the Lower Gorge. Both are popular sites for white water rafting. The Lower Taos Box is a deep, majestic canyon with strong rapids, providing for a medium to very difficult river running experience. The Lower Gorge has less challenging rapids, and rafting opportunities are of only medium difficulty.

## SURVEY DESIGN AND IMPLEMENTATION

This article is based on data collected by mail surveys of random samples of nonresident visitors (i.e., visitors who do not reside in the locality of the site) to the four study sites during spring and summer 1992.[2] The Ramsey Canyon survey sampling frame included bird watchers only. In contrast, the San Pedro survey included visitors participating in all activities. However, 78% of the visitors contacted at the San Pedro RNCA were bird watching, so there is a great deal of consistency in recreational activities across this pair of sites. Both New Mexico surveys focused exclusively on persons engaged in commercial white water rafting. Recreationists were randomly selected for contact at the recreation sites. Ninety percent of the Arizona contacts and 96% of the New Mexico contacts agreed to participate and surveys were mailed within 2 weeks of the on-site contact. All surveys used a payment card format to elicit the recreationists' willingness to pay.[3] (See Cameron and Huppert [3] and Jordan and Elnagheeb [8, 9] for recent comparisons of payment card and referendum (dichotomous choice) CVM formats. In this case, the payment card format is particularly appropriate given that values are being elicited from resource users who have direct experience paying for access to the amenities being valued.) The description of the resource being valued varied

---

[2] Surveys instruments were developed using a focus group process and pre-tested by being mailed to random samples of site visitors and then revised based on the pre-test prior to finalizing the survey design used in the analysis reported here.

[3] The range of values offered on the payment card was based on pre-test results. The true WTP lies between the circled value (the lower bound) and the next higher value (upper bound). This study used the mid-point between these as the bid amount, though more sophisticated payment card estimation procedures have been recommended when payment card intervals are large or the upper most interval includes more than a negligible portion of the sample [3, 8, 9]. The intervals provided on the payment cards in this study were small and less than 1% of the responses lay above the upper most interval. Consequently, the difference between the mid-point used in this study and the log-likelihood function approach used in several other recent studies should be small [3].

TABLE 1

Response Rates

| Survey site | Sent | Received | Response rate | Valid (not protest) zero bids |
|---|---|---|---|---|
| Taos Box | 778 | 590 | 76% | 3 |
| Lower Gorge | 1293 | 970 | 75% | 4 |
| Ramsey Canyon | 454 | 417 | 92% | 37 |
| San Pedro RNCA | 224 | 214 | 95% | 36 |

somewhat between the two Arizona surveys. At one site, the amenity being valued was the continued presence of the Gray Hawk species (and by implication, its riparian habitat). At the other site, the amenity was preservation of a healthy stream-side ecosystem without emphasis on a particular species. In both cases, the underlying resource condition is stated as the preservation of perennial streamflows and associated riparian habitat and wildlife dependent on streamflows. The payment vehicle in both Arizona surveys was a one-time contribution to a non-profit foundation. The two New Mexico surveys elicited the respondents maximum WTP for a rafting trip at the stream flows they actually experienced, with the payment vehicle being additional rafting fees beyond what they actually paid.[4] Follow-up questions elicited motivations for zero bids and served to distinguish valid zero responses from protest zero bids.[5] Table I shows the response rates for all surveys as well as the final sample size. The response rates are relatively high for a mail survey, possibly due to the in-person, on-site contact which preceded the mailing of the surveys.

## BID FUNCTIONS

### *New Mexico Sites*

For the New Mexico data sets we estimated Tobit models for total willingness to pay with a variable truncation point equal to actual trip expenditures, and an exponential heteroskedasticity function. Let $TWP_i$ be the total willingness to pay for the recreation trip, $\mathbf{x}_i$ the $(k \times 1)$-vector of right-hand-side explanatory variables (including a constant term), $\tilde{\mathbf{x}}_i$ the $(k - 1 \times 1)$-vector excluding the constant term, and $p_i$ the actual trip expenditures, respectively, for the $i$th individual at a given site, $b$ a $(k \times 1)$-vector of regression coefficients, $g$ a $(k \times 1)$-vector of parameters in the exponential heteroskedasticity function, and $\varepsilon_i$ a random error term with $E(\varepsilon_i) = 0$ and $E(\varepsilon_i^2) = \sigma_i^2 = \sigma_0^2 e^{2g'\tilde{\mathbf{x}}_i}$. Then the total willingness to pay

---

[4] Theoretical problems with the interpretation of WTP per day or per trip have recently been discussed by Morey [13].

[5] An analysis of those giving protest bids compared to respondents providing valid bids indicated that protest bidders had lower incomes and less river running experience than other respondents, for the New Mexico sites. For the Arizona sites, income did not differ significantly between protest bidders and other respondents, but protest bidders were older and less involved in bird watching activities.

functions for the New Mexico data sets are of the form

$$TWP_i = \begin{cases} \mathbf{x}_i'b + \varepsilon_i & \text{if } \mathbf{x}_i'b + \varepsilon_i > p_i \\ p_i & \text{if } \mathbf{x}_i'b + \varepsilon_i \le p_i \end{cases}. \tag{6}$$

Assuming that the $\varepsilon_i$ across individuals are independent and normally distributed, the mean total willingness to pay is defined by

$$\begin{aligned} E(TWP_i) &= Pr(\varepsilon_i > -\mathbf{x}_i'b + p_i)E(\mathbf{x}_i'b + \varepsilon_i \mid \varepsilon_i > -\mathbf{x}_i'b + p_i) \\ &\quad + Pr(\varepsilon_i \le -\mathbf{x}_i'b + p_i)p_i \\ &= \left[1 - \Phi\left(\frac{-\mathbf{x}_i'b + p_i}{\sigma_i}\right)\right]\mathbf{x}_i'b + \sigma_i\varphi\left(\frac{-\mathbf{x}_i'b + p_i}{\sigma_i}\right) + \Phi\left(\frac{-\mathbf{x}_i'b + p_i}{\sigma_i}\right)p_i \\ &= \Phi\left(\frac{\mathbf{x}_i'b - p_i}{\sigma_i}\right)(\mathbf{x}_i'b - p_i) + \sigma_i\varphi\left(\frac{\mathbf{x}_i'b - p_i}{\sigma_i}\right) + p_i, \end{aligned} \tag{7}$$

where $\varphi(\cdot)$ and $\Phi(\cdot)$ are the standard normal probability function (pdf) and cumulative distribution function (cdf), respectively, $\sigma_i = \sigma_0 e^{g'\tilde{\mathbf{x}}_i}$, and symmetry of the normal probability distribution has been used to obtain the last expression, i.e., $\Phi(z) = 1 - \Phi(-z)$ and $\varphi(z) = \varphi(-z)$. From (7) we find that the expected value of the compensating variation for the $i$th individual, $E(cv_i) \equiv E(TWP_i) - p_i$, is defined by

$$E(cv_i) = \Phi\left(\frac{\mathbf{x}_i'b - p_i}{\sigma_i}\right)(\mathbf{x}_i'b - p_i) + \sigma_i\varphi\left(\frac{\mathbf{x}_i'b - p_i}{\sigma_i}\right). \tag{8}$$

Note that, as argued above in the model section, Eqs. (7) and (8) are observationally equivalent, reflecting the *a priori* restriction that trip expenditure must have a coefficient of negative unity in the compensating variation equation.

There are three sources of nonlinearity in this empirical specification. First, both the standard normal pdf and the cdf are nonlinear functions of the coefficients and explanatory variables. Second, the standard error of the heteroskedastic residual is a nonlinear function of the parameters and explanatory variables. Third, the pdf, cdf, standard error, and regression function interact nonlinearly. Hence, the conclusion of Downing and Ozuna [6] that equality of the regression coefficients is insufficient to establish the equality of benefit measures also apply to this study.

Tables II and III present the variable definitions and estimated regression results for the two New Mexico sites. Only the results with a heteroskedastic error term are presented because likelihood ratio tests strongly rejected homoskedasticity.[6] The results of these tests are Taos Box, $\chi^2(10) = 32.66$, and Lower Gorge, $\chi^2(10) = 251.82$. The critical value at the 0.1% significance level is $\chi^2_{0.001}(10) = 29.6$.

---

[6] We are grateful to a reviewer for suggesting the heteroskedastic specification.

TABLE II

Variables Used in Bid Functions for Northern New Mexico Data

| Variable | Definition |
|---|---|
| WILLPAY | WTP for flow level experienced in terms of additional river running expenses |
| RAFTEXP | River running expenses per person |
| INCOME | Annual household income (before taxes) |
| LNFLOW | Natural logarithm of the flow level experienced |
| DRLNFLOW | LNFLOW $*$ Dummy variable, which is $= 1$ if respondent goes rafting in northern New Mexico twice or more in a typical year, and $= 0$ otherwise. |
| MAINREAS | $= 1$ if river running was not an important reason for traveling to northern New Mexico<br>$= 2$ if river running was one of several important reasons<br>$= 3$ if river running was most important reason<br>$= 4$ if river running was only reason for trip |
| MRLNFLOW | LNFLOW $*$ Dummy variable, which $= 1$ if MAINREAS $= 3$ or 4, $= 0$ otherwise |
| GENDER | $= 0$ if female, $= 1$ if male |
| NM | $= 1$ if respondent is a resident of New Mexico; $= 0$ otherwise |
| CA | $= 1$ if respondent is a resident of California $= 0$ otherwise |
| NEIGHBOR | $= 1$ if respondent is a resident of Arizona, Colorado, or Texas;<br>$= 0$ otherwise |
| TOOLOW | $= 1$ if flow level experienced was perceived as too low,<br>$= 0$ otherwise |

In general, all coefficients in both regression equations were of the theoretically expected signs. The value of the commodity (a rafting trip) increases with income. Flow level was significant and positive. Elasticity measures indicate that respondents' TWP is particularly responsive to flow levels on the Taos Box, reflecting the fact that flow levels make more of a safety and quality difference on that stretch of the river. The effect of flow level on total TWP is higher for those respondents for whom rafting was the most important or the only reason for their trip to northern New Mexico. About 37% of the Taos Box rafters surveyed fell into this category, compared to only 8% among Lower Gorge rafters. Those respondents who go rafting in northern New Mexico on a regular basis (DR equals 1 if respondents regularity go rafting at least twice a year or more) also are more responsive to flow levels, a greater familiarity with the implications of flow level for recreation quality. This is consistent with Boyle *et al.*'s findings [1] regarding the effect of respondents' prior experience with the recreational activity being valued, in their case white water rafting in the Grand Canyon. In addition to the actual flow levels, the subjective perceptions of these flow levels also seem to affect total WTP. Rafters who considered the flows that they experienced as being ‘‘too low for safe and enjoyable river running'' were willing to pay less for the trip experienced than other rafters. The total value of the trip experienced is higher for those respondents who considered rafting an important reason for their overall trip to northern New Mexico.

### *Arizona Sites*

For the two Arizona sites, trip expenditure was not collected, as it is not relevant given the one-time lump-sum character of the payment vehicle. The empirical

TABLE III

Tobit Results for New Mexico Data

| | Taos Box | | Lower Gorge | |
|---|---|---|---|---|
| Variable | Coefficient | Standard error | Coefficient | Standard Error |
| *Regression function* | | | | |
| Income | 1.284 | (0.7460) | 1.045 | (0.3182) |
| LNFLOW | 10.87 | (13.63) | −1.691 | (3.177) |
| MRLNFLOW | 2.823 | (2.744) | −0.3665 | (1.608) |
| DRLNFLOW | 1.922 | (2.718) | 2.988 | (1.932) |
| MAINREAS | 0.6476 | (8.780) | 9.000 | (2.726) |
| Gender | 7.823 | (8.097) | 2.636 | (2.700 |
| NM | −5.651 | (10.32) | −10.13 | (3.917) |
| Neighbor | 5.649 | (11.53) | 0.2112 | (3.237) |
| CA | 62.92 | (15.84) | 6.333 | (5.161) |
| TOOLOW | −14.79 | (13.90) | −4.373 | (3.047) |
| Constant | 9.790 | (88.10) | 40.38 | (19.84) |
| *Heteroskedasticity function* | | | | |
| Income | 0.0176 | (0.0060) | 0.0372 | (0.0038) |
| LNFLOW | 0.0463 | (0.1224) | 0.2449 | (0.0537) |
| MRLNFLOW | 0.0226 | (0.0209) | 0.1310 | (0.0231) |
| DRLNFLOW | 0.0302 | (0.0159) | 0.0836 | (0.0295) |
| MAINREAS | 0.2007 | (0.0819) | 0.0177 | (0.0515) |
| Gender | 0.1111 | (0.0670) | 0.0673 | (0.0494) |
| NM | −0.2818 | (0.0881) | −0.7644 | (0.1297) |
| Neighbor | 0.2629 | (0.0793) | −0.0634 | (0.0504) |
| CA | 0.2251 | (0.1906) | 0.1760 | (0.0583) |
| TOOLOW | 0.0468 | (0.0919) | −0.1853 | (0.0836) |
| $\sigma$ | 22.47 | (17.19) | 5.149 | (1.927) |
| Log-likelihood | −2007.12 | | −3128.23 | |

model is specified with the contingent valuation bids for compensating variation as the dependent variable,

$$cv_i = \begin{cases} \mathbf{x}_i' b + \varepsilon_i & \text{if } \mathbf{x}_i' b + \varepsilon_i > 0 \\ 0 & \text{if } \mathbf{x}_i' b + \varepsilon_i \le 0 \end{cases}. \tag{9}$$

Again, assuming that the $\varepsilon_i$ across individuals are independent and normally distributed, the expected value of compensating variation is defined by

$$\begin{aligned} E(cv_i) &= Pr(\varepsilon_i > -\mathbf{x}_i' b) E(\mathbf{x}_i' b + \varepsilon_i \mid \varepsilon_i > -\mathbf{x}_i' b) + Pr(\varepsilon_i \le -\mathbf{x}_i' b) \cdot 0 \\ &= \Phi\left(\frac{\mathbf{x}_i' b}{\sigma_i}\right) \mathbf{x}_i' b + \sigma_i \varphi\left(\frac{\mathbf{x}_i' b}{\sigma_i}\right). \end{aligned} \tag{10}$$

As in the case for the New Mexico data sets, the error term is assumed to be exponentially heteroskedastic, with $\sigma_i = e^{g'\tilde{\mathbf{x}}_i}$. We focus on the heteroskedastic case because the likelihood ratio tests for homoskedasticity strongly rejected this hypothesis. The results of these tests are: Ramsey Canyon, $\chi^2(8) = 85.24$; San Pedro Birdwatchers, $\chi^2(8) = 107.53$; and San Pedro, All Users, $\chi^2(8) = 103.01$. The critical value at the 0.1% significance level is $\chi^2_{0.001}(8) = 26.1$.

TABLE IV

Variables Used in Bid Functions for Southern Arizona Data

| Variable | Definition |
|---|---|
| Income | Household income (before taxes) in previous year |
| INCSQ | INCOME squared |
| LNAGE | Natural logarithm of the age of the respondent |
| Education | Highest year of formal schooling completed |
| TRIPNUM | Number of trips to the site in the past 2 years |
| Foreign | = 1 if respondent is from a foreign country, = 0 otherwise |
| TRIPSQ | TRIPNUM squared |
| MAINREAS | = 1 if visiting the site was the main reason for the trip to the Sierra Vista area (non-residents)/for the day-trip (locals)<br>= 0 otherwise |
| $\sigma$ | Standard error of the estimate in the Tobit regression |

Tables IV and V show the variable definitions and regression results for Ramsey Canyon and San Pedro RNCA. Since the sample of Ramsey Canyon visitors was designed to include only bird watchers, two versions of the San Pedro regressions are presented: (1) restricting the sample of San Pedro respondents to the 77.6% of respondents who indicated that they were bird watching; and (2) using the full sample, i.e., including visitors who were not bird watchers. This variation in samples is used later in the article to observe how benefit transfer performs differently as the site visitor population sampled is varied.

All variables had the expected sign. WTP increases with the number of times respondents visit San Pedro RNCA in a typical year. Since the WTP question was phrased as a one-time contribution, respondents consider their sum of discounted values over future uses and the number of expected future uses is likely to be larger for those respondents with multiple past visits who have formed a pattern of regular visitation. Visitors from foreign countries are willing to pay less than U.S. citizens, possibly reflecting the fact that these visitors are less likely to return to these sites. Respondents for whom visiting the study site was the main reason for coming to southeastern Arizona were willing to pay more than other respondents. Higher levels of education also were associated with higher WTP values.

## BENEFIT TRANSFER EVALUATION PROCEDURE

In general, if the values obtained from benefit transfer are not statistically different from those obtained through site-specific estimation, convergent validity is established. Several tests for convergent validity are possible. For the New Mexico data sets, let the Tobit regression for compensating variation at the study site (subscript $s$) be denoted as

$$cv_{si} = \begin{cases} \mathbf{x}'_{si} b_s - p_{si} + \varepsilon_{si} & \text{if } \mathbf{x}'_{si} b_s - p_{si} + \varepsilon_{si} > 0 \\ 0 & \text{if } \mathbf{x}'_{si} b_s - p_{si} + \varepsilon_{si} \leq 0 \end{cases}, \tag{11}$$

TABLE V

Tobit Results for Southern Arizona Data

| | Ramsey Canyon | | San Pedro RNCA | | | |
|---|---|---|---|---|---|---|
| | | | Birders | | All Users | |
| | Coefficient | Standard error | Coefficient | Standard error | Coefficient | Standard error |
| *Regression function* | | | | | | |
| Income | 13.88 | (6.894) | −0.0872 | (6.569) | −2.732 | (4.906) |
| INCSQ | −0.3405 | (0.2265) | 0.0415 | (0.2646) | 0.1369 | (0.2040) |
| LNAGE | −9.913 | (54.96) | −88.55 | (49.26) | −49.40 | (26.72) |
| EDUCATION | 12.47 | (4.702) | 4.832 | (3.242) | 5.567 | (2.607) |
| TRIPNUM | 31.92 | (15.10) | 0.8265 | (11.99) | 1.538 | (7.725) |
| TRIPSQ | −2.232 | (1.262) | −0.4195 | (1.111) | −0.5415 | (0.6865) |
| Foreign | −69.29 | (26.17) | −17.37 | (17.71) | −13.87 | (14.74) |
| MAINREAS | 37.14 | (24.64) | 60.19 | (31.35) | 69.98 | (27.22) |
| Constant | −169.7 | (260.3) | 328.7 | (225.5) | 164.7 | (121.0) |
| *Heteroskedasticity function* | | | | | | |
| Income | 0.1255 | (0.0216) | −0.0742 | (0.0435) | −0.0506 | (0.0377) |
| INCSQ | −0.0048 | (0.0010) | 0.0025 | (0.0016) | 0.0018 | (0.0015) |
| LNAGE | −0.6159 | (0.0964) | −0.5046 | (0.3089) | −0.5080 | (0.2594) |
| Education | 0.1029 | (0.0159) | 0.0984 | (0.0272) | 0.1013 | (0.0241) |
| TRIPNUM | 0.2411 | (0.0527) | 0.5505 | (0.1808) | 0.4043 | (0.1733) |
| TRIPSQ | −0.0267 | (0.0072) | −0.1996 | (0.0495) | −0.1394 | (0.0559) |
| Foreign | −0.8197 | (0.2510) | −0.9540 | (0.3978) | −0.9155 | (0.2728) |
| MAINREAS | 0.0457 | (0.0873) | 1.290 | (0.1478) | 1.112 | (0.1342) |
| $\sigma$ | 188.9 | (87.91) | 176.6 | (237.1) | 156.5 | (181.1) |
| Log-likelihood | −2039.54 | | −752.097 | | −935.507 | |

and that for the policy site (subscript $p$) be denoted as

$$cv_{pi} = \begin{cases} \mathbf{x}'_{pi} b_p - p_{pi} + \varepsilon_{pi} & \text{if } \mathbf{x}'_{pi} b_p - p_{pi} + \varepsilon_{pi} > 0 \\ 0 & \text{if } \mathbf{x}'_{pi} b_p - p_{pi} + \varepsilon_{pi} \leq 0 \end{cases}. \tag{12}$$

For the Arizona data sets, the Tobit regressions for compensating variation at the study site simplifies to

$$cv_{si} = \begin{cases} \mathbf{x}'_{si} b_s + \varepsilon_{si} & \text{if } \mathbf{x}'_{si} b_s + \varepsilon_{si} > 0 \\ 0 & \text{if } \mathbf{x}'_{si} b_s + \varepsilon_{si} \leq 0 \end{cases}, \tag{11$'$}$$

and for the policy site to

$$cv_{pi} = \begin{cases} \mathbf{x}'_{pi} b_p + \varepsilon_{pi} & \text{if } \mathbf{x}'_{pi} b_p + \varepsilon_{pi} > 0 \\ 0 & \text{if } \mathbf{x}'_{pi} b_p + \varepsilon_{pi} \leq 0 \end{cases}. \tag{12$'$}$$

One potential means to test convergent validity of *benefit function transfer*, adapting the procedure used by Loomis [11] to the case at hand, is to test the hypothesis $H_0$: $b_s = b_p, g_s = g_p, \sigma_{0s} = \sigma_{0p}$ using a likelihood ratio test. The test results are: New Mexico, $\chi^2(22) = 276.73$, with 0.1% critical value of $\chi^2_{0.001}(22) =$

48.3; Arizona, all respondents, $\chi^2(18) = 93.64$, and Arizona, birdwatchers, $\chi^2(18) = 65.32$, with a 0.1% critical value of $\chi^2_{0.001}(18) = 42.3$ for both Arizona tests. Using this approach, we strongly reject equality of the benefit functions across sites.

However, as Downing and Ozuna [6] show, due to nonlinearities in the willingness to pay functions, statistical equality of the estimated model parameters does not necessarily imply statistical equality of the resulting benefit estimates. On the other hand (although there is no *a priori* reason to expect this) it also could be the case that the estimated benefit function parameters are statistically quite different but the resulting predicted benefits are not significantly different in a statistical sense. Therefore, we propose several alternative tests of the reliability of benefit transfer using confidence intervals around the predicted benefit estimates.

Let $cv_{p|p}$ be the expected value of $cv_p$ given the Tobit benefit function for the policy site. For the New Mexico data sets we have

$$cv_{p|p} = \Phi(z_{p|p})\left(\bar{\mathbf{x}}'_p b_p - \bar{p}_p\right) + \sigma_{0p} e^{\bar{\bar{\mathbf{x}}}'_p g_p} \varphi(z_{p|p}), \tag{13}$$

where $\bar{\mathbf{x}}_p$ is a $(k \times 1)$-vector of sample means for the independent variables at the policy site, $\bar{\bar{\mathbf{x}}}_p$ is the corresponding $(k - 1 \times 1)$-vector of sample means excluding the constant term, $\bar{p}_p$ is average trip expenditure, and $z_{p|p} = (\bar{\mathbf{x}}'_p b_p - \bar{p}_p)/(\sigma_{0p} e^{\bar{\bar{x}}'_p g_p})$. For the Arizona data sets we have

$$cv_{p|p} = \Phi(z_{p|p})\bar{\mathbf{x}}'_p b_p + \sigma_{0p} e^{\bar{\bar{\mathbf{x}}}'_p g_p} \varphi(z_{p|p}), \tag{13$'$}$$

where here $z_{p|p} = \bar{\mathbf{x}}'_p b_p/(\sigma_{0p} e^{\bar{\bar{x}}'_p g_p})$. Similarly, let $cv_{p|s}$ be the benefit transfer estimate of $cv_p$ at the policy site given the Tobit benefit function for the study site. Then for the New Mexico data sets we have

$$cv_{p|s} = \Phi(z_{p|s})\left(\bar{\mathbf{x}}'_p b_s - \bar{p}_s\right) + \sigma_{0s} e^{\bar{\bar{\mathbf{x}}}'_p g_s} \varphi(z_{p|s}), \tag{14}$$

with $z_{p|s} = (\bar{x}'_p b_s - \bar{p}_p)/(\sigma_{0s} e^{\bar{\bar{\mathbf{x}}}'_p g_s})$, while for the Arizona data sets we have

$$cv_{p|s} = \Phi(z_{p|s})\bar{\mathbf{x}}'_p b_s + \sigma_{0s} e^{\bar{\bar{\mathbf{x}}}'_p g_s} \varphi(z_{p|s}), \tag{15}$$

with $z_{p|s} = \bar{\mathbf{x}}'_p b_s/(\sigma_{0s} e^{\bar{\bar{\mathbf{x}}}'_p g_s})$.

We can compute asymptotic 95% confidence intervals around the two estimates of compensating variation for the policy site using the delta method and a first-order Taylor series expansion. It can be shown that the asymptotic 95% confidence interval over $cv_{p|p}$ is

$$CI^{0.95}_{p|p} = cv_{p|p} \pm 1.96\sqrt{r'_p S_p r_p}, \tag{16}$$

where

$$\begin{aligned} r_p &= \left[\partial cv_{p|p}/\partial b'_p \;\; \partial cv_{p|p}/\partial g'_p \;\; \partial cv_{p|p}/\partial \sigma_{0p}\right]', \\ \partial cv_{p|p}/\partial b_p &= \Phi(z_{p|p})\bar{\mathbf{x}}_p, \\ \partial cv_{p|p}/\partial g_p &= \sigma_{0p} e^{\bar{\bar{\mathbf{x}}}' g_p} \varphi(z_{p|p})\bar{\bar{\mathbf{x}}}_p, \\ \partial cv_{p|p}/\partial \sigma_{0p} &= e^{\bar{\bar{\mathbf{x}}}' g_p} \varphi(z_{p|p}), \end{aligned} \tag{17}$$

and $S_p$ is the $(2k \times 2k)$ variance-covariance of the parameter estimates, $[b_p', g_p', \sigma_{0p}]'$. These expressions are valid for both New Mexico and Arizona data sets with the appropriate definition for $z_{p|p}$. Similarly, the asymptotic 95% confidence intervals over $cv_{p|s}$, denoted by $CI_{p|s}^{0.95}$, is obtained by replacing $cv_{p|p}$ by $cv_{p|s}$, $[b_p', g_p', \sigma_{0p}]'$ by $[b_s', g_s', \sigma_{0s}]'$, $z_{p|p}$ by $z_{p|s}$, and $\mathbf{\Sigma}_p$ by $\mathbf{\Sigma}_s$ in the above formula, while holding the explanatory variables fixed at $\bar{\mathbf{x}}_p$.[7] Finally, let $\bar{c}\bar{v}_p$ denote the sample mean contingent valuation bid for the policy site observations and let $\bar{c}\bar{v}_s$ denote the sample mean contingent valuation bid for the study site observations. The following tests can now be constructed.

## *Comparison of Predicted Values*

### *Convergent validity of benefit function transfer.*

$$H_0: CV_{p|s} = CV_{p|p}$$

This tests whether the compensating variation estimate for the policy site obtained by transferring the benefit function from the study site is statistically different from the original estimate for the policy site. Two comparisons need to be made:

(a) $CV_{p|s} \in CI_{p|p}$,

i.e., testing whether the estimate from benefit function transfer lies within the confidence interval over the original estimate; and

(b) $CV_{p|p} \in CI_{p|s}$,

i.e., testing whether the original estimate lies within the confidence interval over the benefit transfer estimate. It is possible that the two comparisons yield contradicting results, in which case the result of the hypothesis test is ambiguous.

### *Convergent validity of direct benefit transfer.*

$$H_0: CV_{s|s} = CV_{p|p}$$

This tests the reliability of the common practice of simply using the predicted WTP for the study site as a benefit estimate for the policy site, without adjusting for differences in the independent variables. Again, two comparisons are required:

(a) $CV_{s|s} \in CI_{p|p}$; and

(b) $CV_{p|p} \in CI_{s|s}$.

## *Comparison of Actual Sample Means*

By construction, the Tobit prediction of compensating variation exceeds the actual sample mean, because the Tobit model adjusts for the fact that the mean of the error term is greater than zero for the sub-sample of positive (non-zero) compensating variation bids. It is not clear which of the two values should be

---

[7] See Cooper [4] for a recent discussion of calculating confidence intervals for dichotomous choice contingent valuation data.

considered as the best estimate. For completeness, the actual sample mean compensating variation for the policy site ($CV_p$) is also compared to the benefit transfer estimates. In particular, the convergent validity of benefit function transfer is evaluated by testing

$$H_0 \colon CV_p \in CI_{p|s},$$

and the convergent validity of direct benefit transfer is evaluated by testing

$$H_0 \colon CV_p \in CI_{s|s},$$

In addition to the question of statistical equality of site-specific and benefit transfer estimates of WTP, the percentage errors from benefit transfer are of interest to policy makers (Loomis [11]). Since the results of the hypotheses tests depend on the width of the confidence intervals, percentage errors from benefit transfer can be fairly large even when two estimates are not statistically different.

The notion of percentage errors presumes that the site-specific benefit estimate is a ‘‘true’’ measure of benefits. This assumption is questionable to the extent that CVM itself is still under scrutiny, though the scrutiny focuses more on nonuse (or passive use) values rather than on values held by users familiar with the resource. The following versions of percentage errors are computed and reported, given researchers’ and policy makers’ interest in this concept:

*Percentage error resulting from benefit function transfer.*

(a) Between the compensating variation estimate obtained from benefit function transfer and the predicted site-specific estimate:

$$\left(CV_{p|s} - CV_{p|p}\right) * 100 / CV_{p|p}$$

(b) Between the WTP estimate obtained from benefit function transfer and the actual sample mean for the policy site:

$$\left(CV_{p|s} - \overline{CV}_p\right) * 100 / \overline{CV}_p.$$

*Percentage error resulting from direct benefit transfer.*

(a) Between the site-specific predicted WTP estimates for the study site and the policy site:

$$\left(CV_{s|s} - CV_{p|p}\right) * 100 / CV_{p|p}$$

(b) Between the actual sample means for the study site and the policy site:

$$\left(\overline{CV}_s - \overline{CV}_p\right) * 100 / \overline{CV}_p$$

## RESULTS

Table VI presents the various benefit estimates used in the hypothesis tests, along with sample average CVM bids and 95% confidence intervals. Tables VII and VIII present the results of the hypothesis tests and the percentage errors from benefit transfer, respectively.

TABLE VI

Sample Means, Predicted Compensating Variations, and 95% Confidence Intervals

| Policy site | Study site | $\overline{cv}_p$ | $cv_{p\|p}$ | $CI^{0.95}_{p\|p}$ | $cv_{p\|s}$ | $CI^{0.95}_{p\|s}$ |
|---|---|---|---|---|---|---|
| Taos Box | Lower Gorge | 26.68 | 33.50 | [26.68, 38.32] | 4.31 | [1.33, 7.28] |
| Lower Gorge | Taos Box | 20.22 | 21.40 | [15.33, 27.47] | 66.43 | [57.23, 71.63] |
| Ramsey Canyon | San Pedro, all responses | 125.74 | 139.97 | [113.1, 166.8] | 133.68 | [78.07, 189.3] |
| Ramsey Canyon | San Pedro, birders | 125.74 | 139.97 | [133.1, 166.8] | 136.74 | [105.5, 168.0] |
| San Pedro, all responses | Ramsey Canyon | 80.41 | 83.03 | [43.33, 122.7] | 112.26 | [87.63, 136.9] |
| San Pedro, birders | Ramsey Canyon | 90.14 | 90.58 | [70.92, 110.3] | 113.37 | [88.92, 137.8] |

*Note*. $\bar{c}\bar{v}_p$ is the sample mean compensating variation bid at the policy site. $cv_{p|p}$ is the expected value of compensating variation for the policy site given the benefit function for the policy site. $CI^{0.95}_{p|p}$ is the asymptotic 95% confidence interval for $cv_{p|p}$. $c_{p|s}$ is the expected value of compensating variation for the policy site given the benefit function for the study site. $CI^{0.95}_{p|s}$ is the asymptotic 95% confidence interval for $cv_{p|s}$. Average rafting expenditures at the New Mexico sites: Lower Gorge, \$41.66; Taos Box, \$92.97.

For the New Mexico data sets, 90% (nine of ten) of our tests lead to rejection of convergent validity. The lone failure to reject at the 5% significance level ($\bar{c}\bar{v}_p \in CI^{0.95}_{s|s}$ for Taos Box as the policy site and Lower Gorge as the study site) also leads to rejection of convergent validity at the 10% level of significance. Thus, benefits transfer between these two sites is rejected by all of the test procedures at a 90% level of confidence or higher.

For the Arizona data sets, 55% (11 of 20) of the tests lead to rejection of convergent validity of benefits transfers between the two sites. Two of the failures to reject at the 5% significance level ($cv_{p|p} \in CI^{0.95}_{p|s}$ and $\bar{c}\bar{v}_p \in CI^{0.95}_{p|s}$ for San Pedro birdwatchers as the policy site and Ramsey Canyon as the study sight) become rejections at the 10% level. Hence, transferring benefits *from* Ramsey Canyon *to* San Pedro birdwatchers is rejected on all counts at the 90% (or higher) level of confidence. In contrast, there is only a 40% (4 of 10) rejection rate for benefits transfers in the other direction, that is, transferring benefit estimates *from* the San Pedro NRCA *to* Ramsey Canyon.

TABLE VII

Results of Hypothesis Tests Regarding the Convergent Validity of Benefit Transfers

| Policy site:<br>Study site: | Taos Box<br>Lower Gorge | Lower Gorge<br>Taos Box | Ramsey Canyon<br>San Pedro, all responses | Ramsey Canyon<br>San Pedro, birders | San Pedro, all responses<br>Ramsey Canyon | San Pedro, birders<br>Ramsey Canyon |
|---|---|---|---|---|---|---|
| $cv_{p\|s} \in CI^{0.95}_{p\|p}$ | No | No | Yes | Yes | Yes | No |
| $cv_{p\|p} \in CI^{0.95}_{p\|s}$ | No | No | Yes | Yes | No | Yes |
| $cv_{s\|s} \in CI^{0.95}_{p\|p}$ | No | No | No | No | No | No |
| $\overline{cv}_p \in CI^{0.95}_{p\|s}$ | No | No | Yes | Yes | No | Yes |
| $\overline{cv}_p \in CI^{0.95}_{s\|s}$ | Yes | No | No | No | No | No |

TABLE VIII

Percentage Errors From Benefit Transfers

| Policy site: | Taos Box | Lower Gorge | Ramsey Canyon | Ramsey Canyon | San Pedro, all responses | San Pedro, birders |
|---|---|---|---|---|---|---|
| Study site: | Lower Gorge | Taos Box | San Pedro, all responses | San Pedro, birders | Ramsey Canyon | Ramsey Canyon |
| $cv_{p\|p}$ vs $cv_{p\|s}$ | −87.1% | +210.4% | −4.5% | −2.3% | +35.2% | +25.2% |
| $\overline{cv}_p$ vs $cv_{p\|s}$ | −83.8% | +228.5% | +6.3% | +8.8% | +39.6% | +25.8% |
| $cv_{p\|p}$ vs $cv_{s\|s}$ | −36.1% | +56.5% | −40.7% | −35.3% | +68.6% | +54.5% |
| $\overline{cv}_p \in \overline{cv}_s$ | −24.2% | +31.9% | −36.1% | −28.3% | +56.4% | +39.5% |

If, in fact, compensating variation for watching rare birds in the San Pedro NRCA were *the same* as the compensating variation for birdwatching in Ramsey Canyon, then benefits should be just as transferable *from* Ramsey Canyon *to* the San Pedro site as the other way around. However, our empirical results lead us to reject this direction of benefits transfer in 9 out of 10 hypothesis tests. Benefit transfer clearly performs better in one direction than the other, as indicated by a comparison of the percentage errors in Table VIII.

## CONCLUSIONS

In evaluating the convergent validity of benefit transfer, the present study provides two specific examples of benefit transfer performance. While no broad generalizations are possible from two pairs of cases, the following conclusions are posited.

Minor differences in the description of the resource to be valued and in visitor activities do not necessarily cause significant differences in benefit estimates as long as the underlying resource condition is the same, as demonstrated by the Ramsey Canyon and San Pedro RNCA data. Percentage errors for benefit transfer involving these sites ranged from 2 to 54% when only bird watchers were considered. Changes in the recreational focus of site visitors between a study site and a policy site (inclusion of visitors that came for recreational activities other than birding) did increase the percentage errors from a benefit transfer somewhat (4 to 68%; see Table VIII). These types of seemingly minor differences in the object of valuation, visitor activities and site visitor populations are present in many potential applications of benefit transfer.

Measurement of a presumed dominant indicator of site quality (i.e., streamflows for rafting) is not sufficient to assure convergent validity in benefit transfer. Other, more difficult to quantify, differences in site characteristics can influence the quality of recreation experiences and apparently can cause substantial biases in benefit transfer estimates, as the results for the two New Mexico sites demonstrate. In the case of whitewater rafting, flow levels are not adequate measures of site characteristics for inter-site comparisons because similar flow levels do not necessarily provide comparable rafting experiences across different rivers or across different stretches of the same river. In this study, benefit transfer did not perform well between the faster, narrow, deep canyon Taos Box section of the Rio Grande and the wider, slower Lower Gorge—even though these are adjacent reaches of the same river.[8] Taos Box trips also are longer (on average) than Lower Gorge trips and this seemingly minor difference in the amenity being valued no doubt contributes somewhat to the poor performance of benefit transfer. Such differences in seemingly similar amenities are commonly encountered in public agency benefit transfer contexts.

Differences in the market for the good, especially the availability and price of substitutes, if not accounted for, also are likely to cause biases in benefit transfer estimates. Multi-site studies would be helpful in determining the effect of various site characteristics and market conditions on the performance of benefit transfers. Pooled regressions for the New Mexico data, including the stretch of the river as a variable, showed that rafters on the Taos Box are on average willing to pay $48.98 more for a rafting trip ($t$-statistic = 14.2) than those rafting the Lower Gorge. In order to determine which site characteristics (e.g., scenic attributes of riverside corridor, trip length, classification of rapids) contribute to this difference in WTP, and to what extend, more inter-site studies are required.

The analysis presented in this study led to a rejection of the validity of direct benefit transfer, that is, the transfer of a mean site benefit estimate. The direct transfer of the study benefit site estimate to another site led to large percentage errors, even under the controlled conditions of this study, which seemed, *a priori*, well suited to benefit transfer. The practice of applying direct benefit transfer between sites which differ far more in quality, location, visitor characteristics, availability of substitutes, and object of valuation than the ones described in this study is, therefore, questionable. If benefit transfer is going to be used in policy analysis, benefit function transfer is preferable wherever possible. Benefit function transfer performed better in nearly all of the cases examined here than did the direct transfer of the site estimate, consistent with the findings of Loomis [11].

This study did not unequivocally reject the hypothesis that benefit function transfer between Ramsey Canyon and the San Pedro RNCA yields valid estimates. This is not surprising given that the two sites are very similar with respect to location, visitor characteristics and quality of the recreational experience. The Arizona results suggest that minor differences in the phrasing of the CVM question do not necessarily lead to substantial biases in benefit transfer estimates. However, the fact that benefit function transfer performed better in one direction (i.e., Ramsey Canyon as the policy site and San Pedro as the study site) than when the policy site and study site are reversed is troubling.

The validity of benefit function transfer was rejected for the Rio Grande data, based on the hypothesis tests developed. The benefit transfer analysis demonstrated large statistical differences between the site-specific and the benefit transfer estimates. From these results, we infer that the information commonly incorporated in bid functions is not sufficient for benefit function transfer when the market conditions for the amenity being valued vary across sites, i.e., they are not close substitutes for one another. Further, as noted previously, the use of flow level as an indicator of recreation quality does not sufficiently reflect differences in the quality of the whitewater experience at the two river reaches studied. Possibly, benefit function transfer would have performed better had a richer mix of site

---

[8] Though these are adjoining river reaches, the flow level at a given point in time differs between them due to water diversions and to tributary inflows.

quality variables been available for use in the bid function. However, such site-specific data for both the policy and the study site often is absent for the contexts in which public agencies wish to use benefit transfer.

The two Arizona sites involved valuing a hypothetical change in the quality of a natural amenity, specific deteriorations in streamflow and habitat quality from the actually observed by visitors during their site visit. In contrast, the two New Mexico case studies measured total willingness to pay for an amenity at a quality level actually experienced by the respondents (whitewater rafting trip, with no change in the quality of the amenity posited). *A priori*, one would expect New Mexico respondents to be able to give more precise WTP responses since they were not asked to consider a hypothetical change in amenity quality, while the Arizona respondents were asked to consider a hypothetical change. This *a priori* expectation is supported by the differing sizes of the standard errors and, consequently, the confidence intervals between the New Mexico and Arizona sites. The confidence intervals are notably wider for the Arizona data. If CVM measures are indeed more precise and specific to the study site when respondents need not consider a hypothetical change, then this may explain why the (more precise) New Mexico benefit measures did not transfer well to the other site, while (less precise) benefit estimates involving hypothetical changes in amenity quality (Arizona) transferred better to another site.

Clearly, economists need to learn more about the effects of resource quality, market conditions and other possible factors on benefit transfer performance. In the meantime, policy makers need to carefully choose appropriate sites for benefit transfer. In particular, the study and the policy sites should be similar with respect to recreation activities, quality of the recreational experience and the availability of substitutes. The large percentage errors from benefit transfer between the Taos Box and the Lower Gorge suggest that if the study and policy sites are not chosen with adequate care, benefit transfer can produce misleading results.

> Analysts will never know the true form of the process underlying people's behavior. Moreover, complete observation of all the factors influencing people's decisions and their choices is impossible. In other words, all applied models are wrong. What is at issue, then, is how important these errors are to their intended use. (V. K. Smith [15])

While many economists have reservations about the appropriateness of benefit transfer, public agencies currently are using benefit transfer due to budget and time constraints that prevent them from conducting original benefit estimation studies. In order to ensure that economic tools are being applied in an appropriate manner by public agencies, it is desirable for the environmental economics profession to eventually be able to provide policy makers with guidelines on the conditions under which benefit transfer can provide reliable estimates of nonmarket benefits. The results reported here contribute to a small, but growing, body of empirical studies evaluating benefit transfer, a body of work which eventually can lead to development of robust criteria for use of benefit transfer. Our results imply that we ought to be skeptical of many efforts to transfer benefit estimates from one site, resource type, or environmental activity to another.

Ultimately, the intended use of the benefit estimate determines whether benefit transfer is appropriate and provides adequate reliability. The value of the additional information from a site-specific study has to be compared to the cost of obtaining site-specific benefit estimates, a cost that is prohibitive for state and federal agencies dealing with many sites. The largest percentage error in the results reported here was 228.5% (direct benefit transfer involving the New Mexico sites). However, of the 24 comparisons of benefit measures provided in Table VIII, only two involved errors exceeding 100% and 16 out of 24 indicate errors of less than 50%. Benefit estimates involving this magnitude of error may provide useful information to policy makers who are conducting a preliminary assessment of a proposed policy's impacts. However, when precision matters in the intended policy application, the appropriateness of benefit transfer is questionable. This study indicates that direct benefit transfer involving seemingly similar sites can produce notable errors in benefit estimates.

## REFERENCES

1. K. J. Boyle, M. P. Welsh, and R. C. Bishop, The role of question order and respondent experience in contingent valuation studies, *J. Environ. Econom. Management* **25**, 5–80 (1993).
2. K. J. Boyle and J. C. Bergstrom, Benefit transfer studies: Myths, pragmatism, and idealism, *Water Resour. Res.* **28**, 657–663 (1992).
3. T. A. Cameron and D. D. Huppert, OLS versus ML estimation of non-market resource values with payment card interval data, *J. Environ. Econom. Management* **17**, 230–246 (1989).
4. J. C. Cooper, A comparison of approaches to calculating confidence intervals for benefit measures from dichotomous choice contingent valuation surveys, *Land Econom.* **70**, 111–122 (1994).
5. W. H. Desvousges, M. C. Naughton, and G. R. Parsons, Benefit transfer: Conceptual problems in estimating water quality benefits using existing studies, *Water Resour. Res.* **28**, 675–683 (1992).
6. M. Downing and T. Ozuna, Jr., ''Testing the Reliability of the Benefit Function Transfer Approach,'' Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN (1994).
7. J. M. Halstead, B. E. Lindsay, and C. M. Brown, Use of the tobit model in contingent valuation: Experimental evidence from the Pemigewasset Wilderness Area, *J. Environ. Management* **33**, 79–89 (1991).
8. J. L. Jordan and A. E. Elnagheeb, Consequences of using different question formats in contingent valuation, *Land Econom.* **70**, 97–110 (1994a).
9. J. L. Jordan and A. E. Elnagheeb, Differences in contingent valuation estimates from referendum and checklist questions, *J. Agric. Resour. Econom.* **19**, 115–128 (1994b).
10. A. J. Krupnick, Benefit transfers and valuation of environmental improvements, *Resources* Winter (1993).
11. J. Loomis, The evolution of a more rigorous approach to benefit transfer: Benefit function transfer, *Water Resour. Res.* **28**, 701–705 (1992).
12. R. C. Mitchell and R. T. Carson, ''Using Surveys to Value Public Goods: The Contingent Valuation Method,'' Resources for the Future, Washington, DC (1989).
13. E. R. Morey, What is consumer surplus per day of use, when is it a constant independent of the number of days of use, and what does it tell us about consumer's surplus? *J. Environ. Econom. Management* **26**, 257–270 (1994).
14. J. J. Opaluch and M. J. Mazzotta, Fundamental issues in benefit transfer and natural resource damage assessment, *in* ''Benefits Transfer: Procedures, Problems, and Research Needs,'' Workshop Proceedings, Association of Environmental and Resource Economists, Snowbird, UT, June (1992).
15. V. K. Smith, On separating defensible benefit transfers from smoke and mirrors, *Water Resour. Res.* **28**, 685–694 (1992).